Maximizing Intra-individual Correlations for Face Recognition Across Pose Differences

Annan Li[1,2], Shiguang Shan[1], Xilin Chen[1] and Wen Gao[3,1]

[1]Key Lab of Intelligent Information Processing of CAS,
Institute of Computing Technology, CAS, Beijing 100190,China
[2]Graduate University of Chinese Academy of Sciences, 100190,Beijing,China
[3]Institute of Digital Media, Peking University, Beijing,100871,China
{anli,sgshan,xlchen,wgao}@jdl.ac.cn

Abstract

The variations of pose lead to significant performance decline in face recognition systems, which is a bottleneck in face recognition. A key problem is how to measure the similarity between two image vectors of unequal length that viewed from different pose. In this paper, we propose a novel approach for pose robust face recognition, in which the similarity is measured by correlations in a media subspace between different poses on patch level. The media subspace is constructed by Canonical Correlation Analysis, such that the intra-individual correlations are maximized. Based on the media subspace two recognition approaches are developed. In the first, we transform non-frontal face into frontal for recognition. And in the second, we perform recognition in the media subspace with probabilistic modeling. The experimental results on FERET database demonstrate the efficiency of our approach.

1. Introduction

As one of the most active research topic, automatic face recognition has received significant attention in computer vision and pattern recognition. After more than 30 years of research, high performance can now be achieved under controlled conditions. But when variations due to extrinsic factors like pose, illumination and expression are present the performance drops significantly[18]. The problem of face recognition is far from being solved, for these variations are very common in real-world applications. Of these, pose change is one of the most important and difficult issues for face recognition. In this paper, we focus on the most common and challenging scenario in which only a single enrolled image is available for each person and the probe image is taken from a different viewpoint.

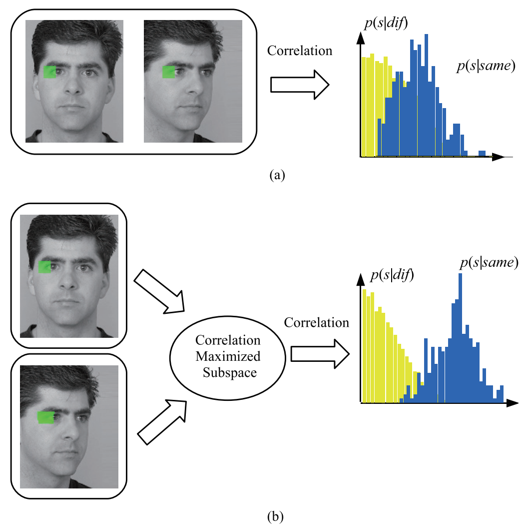

Figure 1. Illustration of how maximizing intra-individual correlations leads to viewpoint invariance. In the above figure, s denotes the correlation value between a pair of corresponding patches. $p(s|same)$ and $p(s|dif)$ are the distributions of the same and different identity respectively. (a) Pose variations confuse the distributions. (b) In the correlation-maximized subspace, the distributions become better separated.

Variations in face appearance due to pose are related to two factors: the viewpoint and the 3D facial shape. In the 2D image taken from a different viewpoint, the locations of surface points on face change differently by reason of 3D facial shape. That is to say, a pair of closer points in one pose may become far away from each other in another pose. This inner structure distortion of image leads to the difficulty of alignment. Furthermore, the 3D shape of human face is not ideally convex. Its concavity can cause occlusion, viz. when pose changes, some visible parts on face may become invisible while some invisible parts may become visible.It leads to a special phenomenon in the pose problem that the distance between two different individual with similar pose is smaller than the distance between the same individual under different pose. Typical frontal face recognition methods, such as Eigenface[17] and Fisherface[2], usually convert face images directly into equal length vectors without any treatment to the problem of image inner structure distortion and occlusion. Thus, it is not surprising that their performances degrade dramatically when pose variation gets bigger.

For the tight connection between the pose variation and 3D shape, 3D prior information is used to enhance the recognition performance. In this type of approaches, the 3D morphable model proposed by Blanz and Vetter[4] is considered the state of the art. By fitting the statistical 3D model to the input face, high recognition rate can be achieved using the representation coefficients or the transformed images[3]. Although the optimization process of fitting guarantees the reconstruction accuracy, but it also bring the problem of high computational complexity. Liu and Chen[12] proposed a similar approach using a simple 3D ellipsoid instead of complex 3D face model. The simplicity of ellipsoid can reduce the computational cost, but obviously, it also limits the accuracy. Although promising, the 3D geometry based approaches still face some hard obstacles, such as accurate initial feature point alignment.

Besides the 3D geometry based approaches, building statistical models is another popular way to tackle the pose invariant face recognition problem. Hitherto, a typical statistical approach is the eigen light-field method proposed by Gross et al.[6]. They build a complete appearance model including all possible pose variations. A test image can be viewed as a part of this complete model. The missing parts are estimated from the available parts. Recognition is performed by comparing the coefficients of the complete appearance model. To reduce variations between different poses, one approach is to transform the model from one pose to another. Sanderson et al.[16] transform the frontal face model to non-frontal views for extending the gallery set and perform the verification using a Bayesian classifier based on mixtures of Gaussians. Similarly, Lee and Kim [10]transform the non-frontal image to frontal in linear feature space. Recently Prince et al.[15] propose a new algorithm based on learning the tied factors between different view points. Based on these factors, recognition is performed with probabilistic distance metirc modeling.

Since local patches are considered more robust to the pose variations than the holistic appearance, patch based approaches are developed in recent years. Kanade and Yamada[8] propose a patch based approach for pose invariant face recognition using Gaussian probabilistic model and Bayesian classifier. Lucey and Chen[13] extend Kanade and Yamada's approach by modeling the joint appearance of frontal patches and holistic non-frontal images. Recently, Ashraf et al.[1] make a further improvement by learning the patch correspondences based on 2D affine transform. By learning the parameters of affine transform on face images across pose, some 3D geometry information is involved. Different from the foregoing approaches that directly measures similarity between patches, Chai et al.[5] perform linear regression on local patches for virtual frontal view synthesis. Each frontal patch is predicted from the corresponding non-frontal patch, then the final virtual frontal view can be synthesized by overlapping the predicted patches. Finally, recognition is performed on the synthesized images.

A key problem in pose robust face recognition is how to measure the similarity between two vectors with unequal length and inner distortion. For example, for a rectangle patch on frontal face, its corresponding region on non-frontal face will expand or shrink with geometric distortion. Thus, a pair of corresponding vectors viewed from different poses are different in length and inner structure. Previous approaches usually measure the similarity of vectors by directly point-to-point matching, e.g. the sum of the squared differences (SSD) in [8, 12, 13, 1]. This measurement requires aligned vectors with equal length, which is a contradiction for the pose variation. To tackle this problem, we propose a novel approach for pose robust face recognition, in which the similarity is measured by correlations in a media subspace between different poses on patch level. The contributions of this paper include:

- We construct a media subspace between different poses by Canonical Correlation Analysis (CCA) [7]. The intra-individual correlations are maximized in this subspace, such that the similarity of patches between different poses can be well measured by the correlation of their projections in the subspace (see Figure 1).
- Based on the media subspace we develop two recognition approaches, i.e., generating virtual frontal views and correlation based classification in the media subspace with probabilistic modeling.

The rest of the paper is organized as follows. In section 2, we describe how the intra-individual correlations are maximized by CCA and 3D patch correspondences. Two correlation based recognition methods are given in section 3. Section 4 shows experimental results of the proposed method. Finally, we draw conclusions in section 5.

2. Maximizing Intra-individual Correlations

In this section, we describe two strategies for maximizing the intra-individual correlations. In the first, we employ 3D patch correspondences to reduce geometric distortion induced by pose. In the second, we construct a media subspace between frontal and non-frontal poses.

2.1. 3D Patch Correspondence

Previous patch based works[8, 12, 13, 5, 1] proved that local patches on face are more invariant to pose change than the holistic images. However, human face has a complex 3D geometric structure. So, if a frontal face is divided into rectangle patches, it is hard to accurately model their corresponding regions on non-frontal face in similar rectangles. To tackle this problem, Chai et al.[5] use a cylindrical 3D model to reduce horizontal distortion, while Liu and Chen[12] use a 3D ellipsoid to map faces into a texture map. However, due to the complexity of 3D faces, neither the cylindrical model nor the ellipsoid model is good enough to represent 3D facial structure. Recently, Ashraf et al.[1] proposed a method to learn the parameters of 2D affine transform between frontal and non-frontal patches. The warped patches in non-frontal view obtained in their approach can reflect some 3D structure information of face. But these patches are not adjacent to their neighbor patches, and this approach can not deal with the problem of the invisible regions.

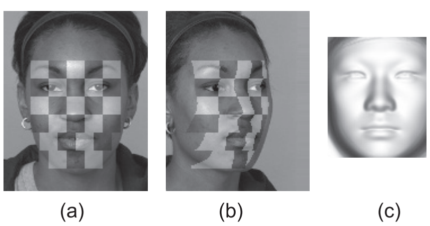

Figure 2. Example of 3D patch correspondences based on a generic 3D face model.(a) Non-overlapping patch division on the frontal face. (b) The corresponding regions on non-frontal face. (c) The generic 3D face model.

In this paper, we find that the patch misalignment can be reduced efficiently by simply using a generic mean 3D face model. By rotating the 3D model, the corresponding region of a frontal patch in non-frontal view can be obtained easily. As illustrated in Figure 2, the corresponding regions can fit the geometric structure of the non-frontal faces pretty well. Compared with the results in [1], the patch correspondences obtained by our method are more close to the real facial shape. The geometric distortion induced by pose can be greatly reduced. Thus, the correlation of each frontal and non-frontal patch pair can be enhanced.

2.2. Constructing the Correlation-Maximized Subspace

As shown in last sub-section, if we vectorize the corresponding patches of different poses, the vectors we get might be different in length due to the visibility or invisibility of some surface points. Thus, the frontal and non-frontal patches form two different subspaces. For measuring the similarity, we use Canonical Correlation Analysis [7] to construct a media subspace between the frontal and non-frontal subspaces.

Let (X, Y) be the training set of vectorized patches on a certain facial region from two different views as we defined in last sub-section, where $X = \{x_1, x_2 \dots, x_n\}$, $Y = \{y_1, y_2 \dots, y_n\}$. Both X and Y are normalized to zero mean. Our Goal is to find two sets of basis vectors, each for one pose, such that the correlations between the projections of variables onto them are mutually maximized. Denote the basis vectors as $W_x = \{w_{x1}, w_{x2} \dots, w_{xk}\}$ and $W_y = \{w_{y1}, w_{y2} \dots, w_{yk}\}$. For a pair of basis vectors (w_x, w_y), the correlation ρ between the projections $w_x^T X$ and $w_y^T Y$ is

$$\rho = \frac{E[w_x^T X Y^T w_y]}{\sqrt{E[w_x^T X X^T w_x] E[w_y^T Y Y^T w_y]}} \tag{1}$$

Here, $E[f(x, y)]$ is the empirical expectation of function $f(x, y)$.

Considering the means of X and Y are zero, the total covariance matrix of (X, Y) can be written as:

$$C_{total} = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} = E\left[\begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}^T\right] \tag{2}$$

where C_{xx} and C_{yy} are the within-pose covariance matrices of X and Y respectively and $C_{xy} = C_{yx}^T$ is the within-individual covariance matrix between two different poses. Thus, the object function can be described as:

$$\rho = \max_{W_x, W_y} \frac{W_x^T C_{xy} W_y}{\sqrt{W_x^T C_{xx} W_x W_y^T C_{yy} W_y}} \tag{3}$$

The solution of W_x and W_y can be found by solving the following eigenvalue equations:

$$\begin{aligned} C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} W_x &= \rho^2 W_x \\ C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} W_y &= \rho^2 W_y \end{aligned} \tag{4}$$

Only one of the equations needs to be solved, because the solutions are related by

$$\begin{aligned} C_{xy} W_y &= \rho \lambda_x C_{xx} W_x \\ C_{yx} W_x &= \rho \lambda_y C_{yy} W_y \end{aligned} \tag{5}$$

where

$$\lambda_x = \lambda_y^{-1} = \sqrt{\frac{W_y^T C_{yy} W_y}{W_x^T C_{xx} W_x}} \tag{6}$$

When W_x and W_y are optimized, the correlation between projection $W_x^T x_i$ and $W_y^T y_i$ is maximized. That is to say, the correlation between a pair of variables that labeled as the same identity is maximized. Thus, we denote the constructed media subspace as intra-individual correlation-maximized subspace.

Additionally,when the unequal-length vectors of different view are projected into the media subspace using the basis vectors respectively, the length of the projected vectors will become equal. Therefore, the problem of measuring similarity between unequal-length vectors can be solved. How maximizing intra-individual correlations leads to viewpoint invariance is illustrated in Figure 1.

3. Recognition Algorithms

Based on the correlation-maximized subspaces, two recognition algorithms are developed, i.e., virtual frontal view synthesis and correlation based classification with probabilistic modeling. As illustrated in Figure 3, the key difference is that the former approach performs recognition in the frontal face subspace, while the latter approach performs recognition in the correlation-maximized subspace.

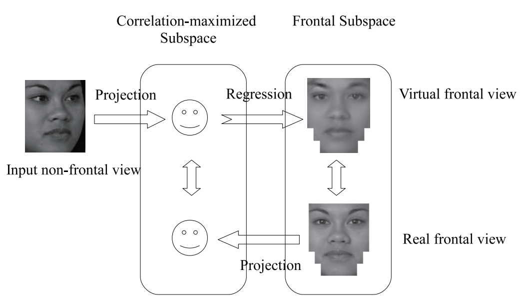

Figure 3. Illustration of two recognition algorithms. In the virtual frontal view based algorithm recognition is performed in frontal face subspace, while in the correlation based classification recognition is performed in the correlation-maximized subspace

3.1. Generating Virtual Frontal Views

Generating virtual frontal views is a popular way to tackle pose invariant face recognition, which is served as a pre-processing phase for the recognition task. Due to the pose variations different poses form different subspaces. Virtual frontal views synthesis can be formulated as a regression procedure that transforms the face from the non-frontal subspace to frontal subspace. An example work of this approach is proposed by Chai et al.[5], in which linear regression is applied. The essence of linear regression is coefficients sharing. That is to say, it uses the best-fit coefficients in non-frontal subspace to reconstruct faces in the frontal subspace. But for two different subspaces, the best-fit coefficients in one subspace are not necessarily the best in another.

Based on the correlation-maximized subspace described in previous section, we can build a bridge between the frontal and non-frontal subspaces. With the constraints derived from both subspaces the coefficients mismatch can be reduced. As shown in Figure 3, at the first step, we project the non-frontal patches into the correlation-maximized subspace. And then, regress them into the frontal face subspace. Similar to [11], we use ridge regression for trade off between accuracy and generalization. The virtual view generating process can be summarized as:

$$y_{virtual} = y_{mean} + R W_x^T (x_{input} - x_{mean}) \tag{7}$$

where x_{input} and $y_{virtual}$ are the input non-frontal and the virtual frontal patch respectively. W_x is the optimized basis, and the R is the regression function:

$$R = Y\hat{X}(\hat{X}\hat{X}^T + \lambda I)^{-1} \tag{8}$$

Here, $\hat{X} = W_x^T X$ is the projection of non-frontal patches, λ is the control parameter of ridge regression.

To smooth the blocking effect, patches are sampled with overlapping. The intensities of overlapped pixels are calculated as the mean of overlapping. Using the virtual frontal images, frontal images based recognition algorithms can be adopted for further recognition.

3.2. Correlation Based Classification with Probabilistic Modeling

Besides the virtual view synthesis, recognition can be conducted directly using the correlation of patches. At first, frontal and non-frontal patches are projected into the correlation-maximized subspace the using the basis vectors W_x and W_y.

$$\begin{aligned} \hat{x_i} &= W_x^T (x_i - x_{mean}) \\ \hat{y_i} &= W_y^T (y_i - y_{mean}) \end{aligned} \tag{9}$$

Here, (x_i, y_i) is the *i-th* patch pair. Then the similarities of corresponding patch pairs are measured by correlation.

$$s_i = \frac{\hat{x_i} \cdot \hat{y_i}}{\|\hat{x_i}\| \|\hat{y_i}\|} \tag{10}$$

Recognition can then be conducted directly by comparing the sum of correlations values of all patches. However, since different patches have different discriminating powers. It is not reasonable to treat them equally. Considering this point, Kanade and Yamada[8] proposed a probabilistic framework for modeling the discriminating power of patches. We combine this framework with the patch correlations. The modeling process is illustrated in Figure 4, and formally described as follows.

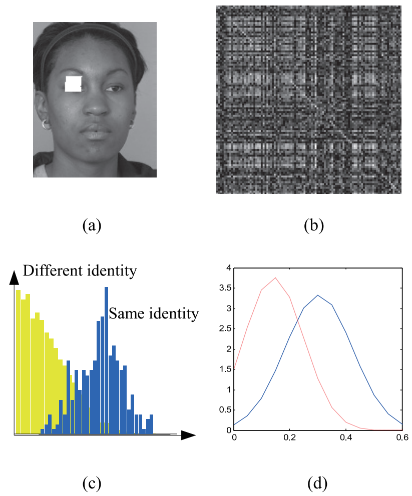

Figure 4. The probabilistic modeling of correlations (a)Example patch (b)The correlation values, diagonal elements are the correlations of same identities. (c)Histograms of correlations: same identity and different identity. (d)The Gaussian fits of (c)

For the i-th patch pair, the conditional probability density that they belong to the same identity is denoted as $P(s_i|same, \phi_p)$. Here, si is the similarity and ϕ_p is the probe viewpoint. Likewise $P(s_i|dif, \phi_p)$ denotes the conditional probability density that of different identities. These distributions are approximated by a Gaussian distribution. Accordingly,

$$P(s_i|same, \phi_p) = \frac{1}{\sqrt{2\pi}\sigma_i^{same}} \exp[-\frac{1}{2}(\frac{s_i - \mu_i^{same}}{\sigma_i^{same}})^2]$$
$$P(s_i|dif, \phi_p) = \frac{1}{\sqrt{2\pi}\sigma_i^{dif}} \exp[-\frac{1}{2}(\frac{s_i - \mu_i^{dif}}{\sigma_i^{dif}})^2] \tag{11}$$

where μ^{same} and μ^{dif}, σ^{same} and σ^{dif} are the means and standard deviations respectively. Based on Bayes rule, the posteriori probability that a patch pair belong to the same identity is

$$P(same|s_i, \phi_p) = \frac{P(s_i|same, \phi_p)P(same)}{P(s_i|same, \phi_p)P(same) + P(s_i|dif, \phi_p)P(dif)} \tag{12}$$

$P(same)$ and $P(dif)$ are the priori probability of the same identity and that of different identity respectively. Based on the sum rule of combining classifiers[9, 8], the total similarity between a gallery image and the probe image can be computed as the sum of the probability values.

$$S(same|I_g; I_p) = \sum_{i=1}^{k} P(same|s_i, \phi_p) \tag{13}$$

where k is the total number of patches.In the classification step, the clamant probe is identified to the gallery identity with highest S value.

4. Experiments

Experiments are performed on the multi-pose subset of FERET database[14]. The FERET database consists of images of 200 subjects captured at 9 different viewpoints. Subsets of each viewpoint are denoted as $ba, bb, bc, bd, be, bf, bg, bh$ and bi, which roughly refer to viewpoint angle of 0°, 60°, 40°, 25°, 15°, -15°, -25°, -40°, -60° respectively. We randomly select 100 subjects for training, while the remaining subjects are used for testing. Images are normalized according to three manually labeled points, i.e., the centers of mouth and two eyes respectively. The frontal face region is divided into patches each of size 16×16. To keep the smoothness, patches are sampled with overlapping. The overlapping step is set to 4 pixels for balance between smoothness and computational cost. Thus, we have 309 patches totally. Figure 5 illustrate the non-overlapped division of 27 patches and the overlapping effect.

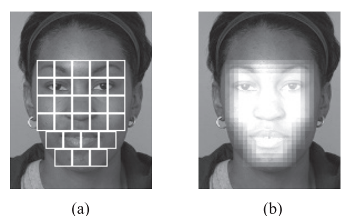

Figure 5. The patch dividing and overlapping. (a) Non-overlapped division of 27 patches. (b) The overlapping effect of 309 patches

4.1. Virtual Frontal View Based Recognition

As described in previous section, frontal faces can be synthesized using patches transformed from non-frontal faces. In Figure 6, example results of virtual frontal view synthesis are shown. For comparison, the results of linear regression on the 3D corresponding patches are also given. We can see that, the proposed synthesis approach is more immune to geometric distortion.

When virtual frontal views are synthesized, recognition can be performed using any frontal face recognition algorithm. In this paper we choose the Fisherface[2] method for its effectiveness and popularity in face recognition. In the FERET database there is only one image of frontal view for each subject. However, the Fisherface method requires at

Figure 6. Example results of virtual frontal view synthesis. (a) Input non-frontal image. (b) Virtual views generated by the proposed method. (c) Virtual views generated by linear regression. (d) The ground truth frontal view

least two samples per subject. We solve this problem by expanding the training set with virtual frontal views obtained from cross validation. We divide the training set into five subsets. For each subset, the non-frontal images are transformed to frontal using the regression function trained on the remaining four subsets. Then we can get a virtual frontal view for each subject. Generating the virtual samples for all poses, we can expand the training set to 9 samples per subject, i.e., one real sample and 8 virtual samples respectively.

We compare the recognition performance using three different kinds of images, i.e., the virtual views synthesized by our approach, those obtained through linear regression and the original images respectively. As shown in Figure 7, under the viewpoints close to frontal, the performances on all kinds of images are close. But when pose angle become larger, our approach performs noticeably better.

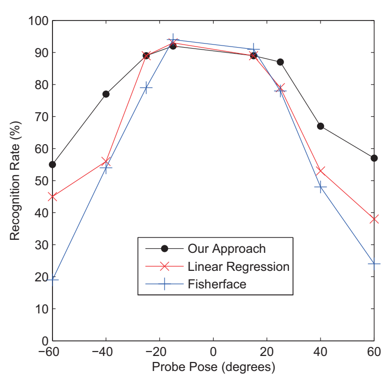

Figure 7. Comparison of recognition performance on virtual frontal view synthesis.

4.2. Correlation Based Recognition

Besides the virtual frontal view synthesis, recognition can also be performed using the correlations. But, when conducting the recognition, we meet similar problem in the virtual frontal views based approach. We have to learn both the correlation-maximized subspace and the prior distribution of correlations on the same training set. To avoid this problem, similar cross validation procedure is performed to learn the prior distributions of correlations. The training set is divided into 5 subsets. In each subset, the vectorized patches are projected into the correlation-maximized subspace that learned from the remaining 4 subsets. Then we get a correlation value for each patch pair in the training set. Hereafter, the prior distribution of same identities and different identities are approximated using these correlation values.

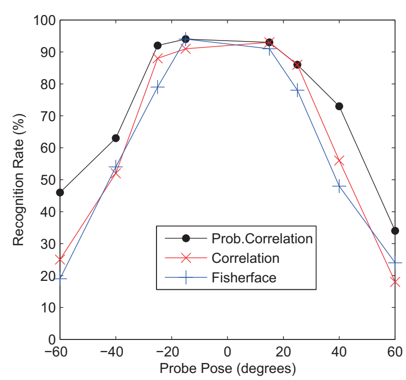

Figure 8. Performance of recognition using probabilistic correlation and correlation and fisherface

In the recognition experiments we set P(Same) = P(Dif) = 0.5. Since there is no prior knowledge, we assume that the probability is equal. To validate the proposed method, we compare the probabilistic approach with the non-probabilistic approach. As shown in Figure 8, the probabilistic approach outperforms the non-probabilistic approach. The probabilistic modeling on correlations can improve the performance. Moreover, if we treat the holistic image as a "bigger patch", and build similar probabilistic model on them, the performance in large pose angle can be considerably improved. The results are shown in Figure 9. When viewpoint angle becomes larger, the corresponding regions of frontal patch become smaller and smaller on non-frontal face. The information in these patches is declined, so that the information between frontal and non-frontal patch pairs is unbalanced. Using the holistic images or enlarging the corresponding regions in non-frontal images can reduce this information declination. This may explain the performance promotion induced by using holistic images.

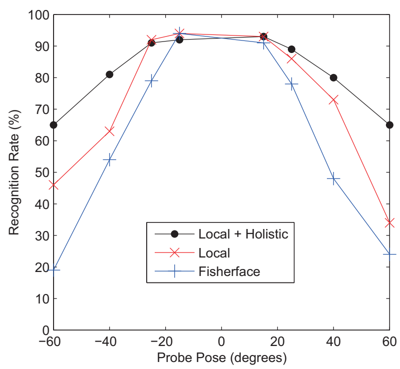

Figure 9. Comparison of performance between the approach using only local patches and that combining with holistic images

5. Conclusions

In this paper, we explore the problem of similarity measurement in pose robust face recognition. A novel measurement approach is proposed, in which the similarity of patches between different poses is measured by correlations in a media subspace constructed by Canonical Correlation Analysis. Based on the media subspace two different recognition algorithms are developed. In the first, we transform non-frontal faces into frontal for recognition, while in the second we perform recognition in the media subspace with probabilistic modeling. Experimental results demonstrate that both of them are efficient. Therefore the proposed similarity measurement is suitable for pose robust face recognition.

A limitation of our approach is the assumption of known pose. Although the head pose can be estimated, it is not clear that how robust to the error of pose estimation our approach is. We will analyze it and combine our approach with pose estimation in the future work.

Acknowledgement

This paper is partially supported by NSFC under contracts No.60332010, No.60728203, No.U0835005; National Basic Research Program of China (973 Program) under contract 2009CB320902; Hi-Tech Research and Development Program of China under contract No.2007AA01Z163; 100 Talents Program of CAS; and ISVISION Technology Co. Ltd.

References

[1] A. B. Ashraf, S. Lucey, and T. Chen. Learning patch correspondences for improved viewpoint invariant face recognition. In *CVPR*, June 2008.

[2] P. N. Belhumeur, P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, 19(7):711–720, 1997.

[3] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter. Face recognition based on frontal views generated from non-frontal images. In *CVPR*, June 2005.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *PAMI*, 25:1063–1074, 2003.

[5] X. Chai, S. Shan, X. Chen, and W. Gao. Locally linear regression for pose-invariant face recognition. *IEEE Trans. on Image Processing*, 16(7):1716–1725, 2007.

[6] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light-fields. *PAMI*, 26(4):449–465, 2004.

[7] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[8] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach toward pose-invariant face recognition. In *CIRA*, July.

[9] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *PAMI*, 20(3):226–239, 1998.

[10] H.-S. Lee and D. Kim. Generating frontal view face image for pose invariant face recognition. *Pattern Recognition Letters*, 27(7):747–754, 2006.

[11] Z. Lei, Q. Bai, R. He, and S. Z. Li. Face shape recovery from a single image using cca mapping between tensor spaces. In *CVPR*, June 2008.

[12] X. Liu and T. Chen. Pose-robust face recognition using geometry assisted probabilistic modeling. In *CVPR*, June 2005.

[13] S. Lucey and T. Chen. Learning patch dependencies for improved pose mismatched face verification. In *CVPR*, June 2006.

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *PAMI*, 22(10):1090–1104, 2000.

[15] S. J. D. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti. Tied factor analysis for face recognition across large pose differences. *PAMI*, 30(6):970–984, 2008.

[16] C. Sanderson, S. Bengio, and Y. Gao. On transforming statistical models for non-frontal face verification. *Pattern Recognition*, 39:288–302, 2006.

[17] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[18] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.